

**Sven Leuckert. 2019. *Topicalization in Asian Englishes: Forms, Functions, and Frequencies of a Fronting Construction*. London: Routledge. xiv + 221 pp.**

**Reviewed by Gabriela Anidora Brozba\***

This book focuses on topicalization in World Englishes, a phenomenon that refers to the placement in initial position of constituents other than the subject, as in *That man I won't marry*. The volume constitutes a comprehensive study of the characteristics of topicalization in four Asian varieties: Hong Kong English (HKE), Indian English (IndE), Singapore English (SingE), and the Philippines English (PhIE). The author is interested in identifying the factors that influence topicalized structures in Asian Englishes. He is also keen on pinpointing whether such structures share commonalities across the studied varieties and to what extent they resemble the standard varieties (i.e. British or American English). The material available in the *International Corpus of English* offers the ground for potential answers to these objectives.

The book has an introduction, three theoretical chapters (namely chapters 2, 3, and 4), 3 empirical chapter (chapters 5, 6 and 7), and conclusions. Additionally, the author provides the readers with a useful index, alongside a list of figures, a list of tables, a list of abbreviations and references.

Chapter 1, "Introduction" (pp. 1-6), outlines the aims and research questions, as well as the structure of the book.

In chapter 2, "Approaching topicalization" (pp. 7-33), the author reviews the literature on topicalization and the notions of "topic" and "topicalization," which he defines as "the marked fronting of constituents" (p. 7). This definition also accounts for instances which show topicalized constituents other than NPs and reveal information other than "discourse-old" or "evoked." In general, researchers have used the terms "topicalization" and "topic" to refer to language structures often found in first language varieties, whereas studies focusing on non-native varieties of English have taken a more inclusive approach, namely more inclusive of innovative structures (Mesthrie 1997).

Chapter 3, "Topic-prominence in Asian contact languages" (pp. 34-61), constitutes a review of the typology of the substrate languages which enrich the pool of linguistic options involved in the development of the four Asian Englishes (in the sense put forward by Mufwene 2001). The author follows Li & Thompson's (1976) criteria to classify languages into "topic-prominent" (e.g. Chinese languages – based on the topic-comment principle), and/or "subject-prominent" (e.g. Indo-European languages – based on the subject-predicate principle). He has, therefore, the task of looking at the ten relevant substrates in the Asian Englishes selected for the study, with the aim of testing the contact hypothesis that topic-prominent substrates increase the frequency of topicalized structures. This undertaking is based mainly on the grammars available for the substrate languages under scrutiny. Leuckert aims to demonstrate that Mandarin and Cantonese are the only clearly topic-prominent languages and would generate a higher frequency of topicalized structures. The other substrates (i.e. Hindi, Bangla, Marathi, Tamil, Telugu, Malay, and Kannada) are found as being both subject-prominent and topic-prominent languages. Tagalog is assigned uncertain status.

Chapter 4, "Development and variety status of four Asian Englishes" (pp. 62-80), investigates the "degree of development" and "variety status" (pp. 62-80), which legitimate the conditions of emergence of the different varieties and their current status in each country. Leuckert discards the traditional Three Circles model of Kachru (1985), which he finds static in nature. He

---

\* University of Bucharest, Department of English, gabriela.brozba@lls.unibuc.ro.

also does away with the more recent model of the Extra- and Intra-Territorial Forces (Buschfeld & Kautzsch 2016), because of its failure to cover postcolonial contexts. The author opts for Schneider's Dynamic Model (2007) for his current study because of its evolutionary view on the postcolonial varieties of English. On the basis of the chosen model, he categorizes SingE as the only variety having reached endonormative status<sup>1</sup>, while the other three (HKE, IndE, and PhIE) are still at the nativization phase, which already involves the emergence of local linguistic features. He also looks at the degree of contact with local varieties, which allows for a continuum from IndE (the "oldest" variety) to HKE (the "youngest" variety).

Chapter 5, "Corpus analysis: Data basis and methodology" (pp. 81-97), is a methodological one. The author provides the source of the data and the procedure followed in searching for relevant cases of topicalization. He also explains the relevance of the variables used to code the examples (i.e. those of a grammatical nature and those of a discourse nature).

Chapters 6 and 7 represent the heart of this empirical work. Chapter 6, "Forms, functions, and frequencies of topicalization" (pp. 98-171), focuses on the analysis of the data from each corpus, by combining quantitative and qualitative techniques. His results of the comparative approach reveal the following: (i) topicalized structures exhibit a variety of forms other than NPs, and (ii) topics are usually evoked, but also new or unused. Chapter 7, "Explaining topicalization frequencies" (pp. 172-193), tackles the major and minor factors involved in explaining the frequency of topicalization. Leuckert aims to show that topicalization is a frequent phenomenon in some varieties, such as IndE or SingE, but far less frequent in others, like HKE or PhIE. The phenomenon of topicalization allows Leuckert to show that Asian Englishes in particular (and World Englishes in general) are shaped by language contact effects, second-language acquisition processes, and social factors. The influence of the substrate language, alongside the development status of the variety at issue may be decisive factors for the frequency of topicalization in SingE. HKE is a younger variety and still exonormative so this can explain the lower frequencies of topicalized structures. PhIE is left without a clear explanation, given the uncertain substrate status of Tagalog and the fact that it has as a superstrate American English (which is not classified typologically).

In chapter 8, "Conclusion and outlook" (pp. 194-197), the author summarizes the findings, takes a new look at the limitations of the study and also identifies topics for further research.

In conclusion, the book on topicalization written by Leuckert is an invaluable addition to the literature on topicalization in general and the one in non-native Englishes in particular. The book has a clear structure, the objectives and research questions are clarified from the beginning. The material used allows for ample comparative analyses, both quantitatively and qualitatively.

The author's cautionary and inclusive tone is appreciated throughout the study. His approach to topicalization in World Englishes is a less restrictive one (e.g. he includes examples such as *Left how many more month* (p. 149) instead of the expected *How many more months are left?*) His multilayered analysis takes into consideration factors such as: language contact effects, second language acquisition, socio-cultural factors, diachronic phenomena and so on. The discussion about problematic / ambiguous examples in chapter 5 is proof of cautionary approach again, especially since some examples were subject to personal interpretation. When analyzing language structures, it can be a daunting task to assign it to different categories, when it comes to the variable "discourse function", which makes it difficult to discriminate between variants, especially in the absence of recordings.

One will also appreciate the author's attention to detail and his diligence when handling challenging categorizations. One such example is the analysis of adverbials (a functionally vague

---

<sup>1</sup> Not surprisingly, Deterding (2007) claims it is the only Asian country in which one can claim English has truly official status and it has been infused in all areas of life.

category), which are not always easy to classify as topicalized constituents (p. 139), especially without access to intonational information. Categorization is also a challenge when it comes to the classification of topics according to discourse function (as mentioned earlier), within which there are four categories: emphasis, contrast, topic continuity, and topic shifting (pp. 158-159).

The book has little that one can find fault with, precisely because of the author's cautionary tone and straightforwardness in going about the personal interpretations of some discourse patterns. Therefore, minor shortcomings such as the lack of information for the superstrate of PhLE and the debatable categorization of some examples are truly insignificant in the bigger picture of a truly remarkable endeavour, which has undoubtedly enriched the knowledge on the behaviour of topicalized structures in non-native Englishes.

### References

- Buschfeld, S. & Kautzsch, A. 2016. Towards an integrated approach to postcolonial and non-postcolonial Englishes. *World Englishes* 36 (1): 104-126.
- Deterding, D. 2007. *Singapore English*. Edinburgh: Edinburgh University Press.
- Kachru, B. B. 1985. Standards, codification and sociolinguistic realism: The English language in the outer circle. In R. Quirk & H. G. Widdowson (eds.), *English in the World: Teaching and Learning the Language and Literatures*, 11-30. Cambridge: Cambridge University Press.
- Li, Charles N. & Thompson, S. A. 1976. Subject and topic: A new typology of language. In Charles N. Li (ed.), *Subject and Topic*, 457-489. New York: Academic Press.
- Mesthrie, R. 1997. A sociolinguistic study of topicalisation phenomena in South African Black English. In E. W. Schneider, (ed.), *English around the World*, vol. 2, *Caribbean, Africa, Asia, Australasia: Studies in Honour of Manfred Görlach*, 119-140. Amsterdam/Philadelphia: John Benjamins.
- Mufwene, S. S. 2001. *The Ecology of Language Evolution*. Cambridge: Cambridge University Press.
- Schneider, E. W. 2007. *Postcolonial English: Varieties around the World*. Cambridge: Cambridge University Press.

**Ursula Stephany & Ayhan Aksu-Koç (eds.). 2021. *Development of Modality in First Language Acquisition*. Berlin: De Gruyter Mouton. x + 593 pp.**

**Reviewed by Veronica Tomescu\***

The volume offers a detailed description of the acquisition of modal verbs, adverbs and other verbal forms with modal value in a number of typologically different languages, some of them less known: German, Lithuanian, Russian, Croatian, French, Romanian, Greek, Estonian, Finnish, Hebrew, Turkish, Korean and two Mayan languages.

Aside from charting the early development of the various linguistic expressions of modality, most papers also cover other topics of interest to the researcher: the influence of socio-economic factors, gender or parenting style on language development, the relationship between child language and child-directed speech (CDS), and the order and timing of emergence of constructions of various cognitive complexity, including the delay in the emergence of epistemic compared to deontic and dynamic contexts.

In "Studying the acquisition of modality: An introduction" (pp. 1-24), the editors state the aims of the volume, sketch its organization, and offer a brief theoretical background.

In the first chapter, "Requests in first language acquisition of German: Evidence from high and low SES families" (pp. 25-78), Katharina Korecky-Kröll investigates the impact of

---

\* University of Bucharest, Department of English, veronica.tomescu@lils.unibuc.ro.

socio-economic status on the use of requests in child and adult German. As also shown in previous research (Hoff-Ginsberg 1991, etc.), parents from lower socio-economic backgrounds prefer a more behaviour-directing parenting style and they use a significantly higher number of requests and especially direct requests compared to parents of higher socio-economic status. The latter group use fewer requests and especially very few direct requests and prefer other speech acts, such as questions for information or statements of facts. Matching these findings, the chapter by Sigal Uziel-Karl on the acquisition of Hebrew modality ("Modality in child Hebrew", pp. 379-420) also reports a reduced number of imperatives in the speech of two high socio-economic status mothers and a higher number of future imperatives, considered less face-threatening.

Overall, the behaviour of the two groups of children in Korecky-Kröll's study mirrors that of their parents, however the differences between the two groups of children are not as significant as between the two groups of adults. Children from both groups produce a similar number of requests and questions for information, since their needs for attention and information are very similar and not influenced by socio-economic background. That being said, it remains true that linguistic development is indeed influenced by social factors. Thus, children of low socio-economic status resort more frequently to imperatives and infinitives with imperative meaning (direct requests in general), on the model of CDS; on the other hand, speech acts expressing permission are more frequent in the utterances of children from higher socio-economic backgrounds; finally, both parents and children of high socio-economic background have a larger inventory of modal verbs and other expressions with modal value.

In "Gender differences in the acquisition of requests in Lithuanian" (pp. 79-112), Viktorija Kavaliauskaitė-Vilkinienė & Ineta Dabašinskienė set out to identify any gender differences reflected in the use and acquisition of requests in Lithuanian child speech and child-directed speech, on the basis of two early longitudinal corpora of a boy and a girl. The study fails to confirm the hypothesis that children's language reflects gender differences in education, but the authors themselves admit the difficulty of drawing any conclusion on the basis of just two corpora, both ending before the third birthday.

Maria D. Voeikova & Kira Bayda, in "Development of directive expressions in Russian child-adult communication" (pp. 113-158), offer a very detailed description, replete with examples, of the various forms expressing requests and prohibitions in adult Russian (in child directed speech) and their development in two longitudinal corpora. The acquisition path is influenced by the frequency of the various forms in the input and consequently by the mothers' different parenting styles: for example, one of the mothers uses numerous hortatives and indirect requests, whereas the other mostly relies on direct requests, and this preference is mirrored in the children's utterances.

Gordana Hržica, Marijan Palmović & Melita Kovacevic, in "Acquisition of modality in Croatian" (pp. 159-190), offer the first systematic description of the early development of agent-oriented modal forms in child Croatian, on the basis of three longitudinal corpora, but the study stops short of the emergence of epistemic uses. Modal verbs are preceded in child language by imperatives and modally used infinitives.

In "Competition of grammatical forms in the expression of directives in early French child speech and child-directed speech" (pp. 191-234), Marianne Kilani-Schoch looks at indicatives and root infinitives used with deontic value in child French and investigates whether they are used in competition with imperatives. In the early stages of acquisition, root infinitives are preferred to imperatives in negative contexts, but no true difference in meaning is observable, due to the limited pragmatic knowledge of the children. The indicative used to express requests appears later than the imperative in child speech and it is not immediately evident that the two forms come to fulfil different functions. Competition between the imperative and the second person singular indicative is observable in child directed speech: the forms are alternately used, sometimes with

similar, but also with differing illocutionary force. The parents try to avoid strong direct prohibitions and negative imperatives are replaced with the indicative or modal verbs. Special attention is given in the chapter to variation sets observable in CDS, the rephrasing of an imperative with an indicative form in the next utterance, or the other way around, with no obvious change in illocutionary force. These variation sets seem to contribute to language development in the child and promote language flexibility.

The chapter “Acquisition of requests in Estonian” by Reili Argus (pp. 315-346) describes the linguistic means employed to express requests in early child Estonian, on the basis of two longitudinal corpora. The order of acquisition is determined by the cognitive complexity of the structure: children start out by producing commands addressed to the interlocutor, followed by hortatives, while indirect requests including a third party as performer are acquired last. The frequency of these various forms in the input is found to have affected the acquisition path to a significant extent.

The impact of CDS on language development is also explored in the chapter by Klaus Laalo, “Directives in Finnish language acquisition” (pp. 347-378), which charts the acquisition path of directives in the early utterances of two Finnish children. Early requests are expressed by means of imperatives, partitives and illatives. Very soon, under the influence of CDS, where these forms are quite frequent, children acquire passive imperatives with hortative meaning and illative forms of the third infinitive. Plural imperatives however, that are rare in CDS, emerge later.

Several papers in the volume explore the cause behind the so-called epistemic gap. It has been documented that epistemic modality emerges later than dynamic and deontic modality in most languages (Hickmann & Bassano 2016), due to the gradual development of metacognitive abilities (Papafragou 1998), although variations in the morphosyntactic realization of modality in different languages may question the validity of this generalization. Epistemic adverbs appear quite early (Avram & Gaidargi, this volume, Stephany, this volume, etc.), and in languages where epistemic modality is (also) realized by means of suffixes, an early emergence of these suffixes has been documented (Choi 1991, etc).

Epistemic modality follows subject-oriented and agent-oriented modality in child Hebrew, in the corpora analysed in Sigal Uziel-Karl’s chapter, “Modality in child Hebrew”, pp. 379-420, although all uses are attested from early on; the two girls in the study employ both modal verbs and (epistemic) modal adverbs.

In “Epistemic modality in Russian child language” (pp. 421-452), Victoria V. Kazakovskaya charts the emergence of epistemic modality in child Russian. The children in the study first learn to convey evaluation of surrounding objects and their properties, later of their own actions and the actions of others, and finally of the mental states of their interlocutors. Uncertainty markers are acquired before certainty markers. This development, together with the precedence that subject- and agent-oriented uses have over epistemic uses, leads the author to propose a theory of mind account for the timing and progression of the acquisition of epistemic modality.

Larisa Avram & Andreea Gaidargi, in the chapter entitled “On the acquisition of dynamic, deontic and epistemic uses of modal verbs in Romanian” (pp. 235-254), investigate 3 longitudinal corpora to document the emergence of modality in Romanian. Subject-oriented dynamic modals are more frequent than and precede deontic uses, but epistemic contexts, with the exception of epistemic adverbs, are absent from the children’s utterances. That epistemic modality emerges late because of gradual cognitive development is contradicted by the presence of epistemic adverbs in early recordings. But the input contains remarkably few epistemic modals, and indeed overall in Romanian epistemic meanings are preferentially expressed by means of adverbs rather than modal verbs. Additionally, the authors propose that the acquisition of epistemic adverbs is more facile than that of the modal verbs which have multiple functions and contexts of use.

Proof for the early production of epistemic contexts is also to be found in the chapter entitled “Development of modality in early Greek language acquisition” by Ursula Stephany (pp. 255-314), which documents in rich detail the emergence of various modalized contexts in child Greek and provides an informative inventory of these contexts in adult modern Greek. In the 5 corpora investigated, epistemic adverbs emerge simultaneously with deontic and dynamic modals. But epistemic uses of modal verbs follow deontic or dynamic contexts and are produced significantly less frequently. All in all, language-specific properties also interfere with the acquisition of modality: whereas the epistemic use of modal verbs seems to follow the emergence of deontic and dynamic contexts in all languages, this order is no longer universally observed in the case of other linguistic realizations of modal verbs, such as modal adverbs.

As for languages where epistemic modality is realized by means of suffixes alongside other types of expressions, such as Turkish and Korean in this volume, early emergence of epistemic contexts has been documented.

Epistemic modality is shown to be acquired before the second birthday in child Turkish in the chapter by Treysi Terziyan & Ayhan Aksu-Koç, “Epistemic and evidential modality in early Turkish child speech” (pp. 453-490). For one of the children, inflectional expressions precede the emergence of epistemic adverbs. However, the multifunctionality of the inflections, which requires the deployment of increased cognitive effort in their use, leads children to prefer epistemic adverbs to express certainty or uncertainty.

“The development of sentence-ending epistemic/evidential markers in young Korean children” by Soonja Choi (pp. 491-524) discusses the order in which epistemic suffixes are acquired in child Korean. The suffixes are acquired quite early, but in a particular order, dictated by cognitive development: first the suffix used to express their own knowledge status, then the suffixes which also incorporate the listener knowledge status, and finally the suffix which conveys information contrasting with the listener’s assessment. Additionally, suffixes which are acquired first are also to be found to a greater extent in the input and have a higher structural resonance, that is the proposition is a partial or full repetition of a prior utterance, significantly simplifying the learner’s task.

The volume concludes with a paper by Barbara Pfeiler & Alejandro Curiel, “The acquisition of evidentiality in two Mayan languages, Yukatek and Tojolabal” (pp. 525-554), describing the expression and acquisition of evidentiality in two Mayan languages, Yukatek and Tojolabal, on the basis of longitudinal corpora, but also narratives in the case of Tojolabal. Like other Mayan languages, Yukatek and Tojolabal have four evidential categories: suffixes marking non-specified sources of information, reported speech, quoted information and information obtained by means of direct observation. In the corpora, quotatives are employed mostly to rephrase the children’s utterances with a view to clarifying them for the researcher, but also in order to prompt the children to talk or otherwise control their behaviour. Quotatives are more frequent in Tojolabal CDS. Reportatives are more frequent than quotatives in the Yukatek CDS and are used to render commands issued by a third party. In the children’s speech, these suffixes emerge early but are scarce. No difference with respect to timing has been observed between the two languages. In both Yukatek and Tojolabal, reportatives emerge later than quotatives possibly due to their being cognitively more challenging.

An exceptionally engaging section of the paper describes a verbal contest where children show off their rhetorical skills by means of narratives. The first child tells a funny story, which is then retold by a second child, in a more sophisticated manner, and so on, in a chain of narratives. The retelling of the others’ narratives necessarily requires the use of quotatives and reportatives, which makes this contest a suitable material for assessing the children’s proficiency in the use of these suffixes. The section includes one such chain of narratives, which, apart from the analysis of the evidential forms, also represents a tiny yet intriguing window into Tojolabal culture.

In the final chapter, “Conclusions” (pp. 555-576), the authors sum up the main theoretical points covered by the volume.

This collection of studies is bound to be useful for all researchers interested in language acquisition and the study of modality. Aside from the valuable theoretical insights, it also provides delightful reading for anyone fascinated by the study of languages in general, especially so because it covers a range of less studied languages, and, more particularly, it offers a rich catalogue of the various modal verbs and other verbal forms with modal value in the languages described here.

### References

- Choi, S. 1991. Early acquisition of epistemic meanings in Korean: A study of sentence-ending suffixes in the spontaneous speech of three children. *First Language* 11: 93-119.
- Hickmann, M. & Bassano, D. 2016. Modality and mood in first language acquisition. In J. Nuyts & Johan van der Auwera (eds.), *The Oxford Handbook of Modality and Mood*, 430-447. Oxford: Oxford University Press.
- Hoff-Ginsberg, E. 1991. Mother-child conversation in different social classes and communicative settings. *Child Development* 62 (4): 782-796.
- Papafragou, A. 1998. The acquisition of modality: Implications for theories of semantic representations. *Mind and Language* 13 (3): 370-399.

**Sean Wallis. 2021. *Statistics in Corpus Linguistics Research: A New Approach*. Abingdon: Routledge. xiii + 355 pp.**

**Reviewed by Mihaela Buzec\***

The book *Statistics in Corpus Linguistics Research* is not just yet another book on statistics. It is an aid for students and researchers alike, who struggle to approach statistics meaningfully on their own. The author, Sean Wallis, is a researcher specialized in corpus linguistics, artificial intelligence, and statistics, who is part of the Survey of English Usage research unit at University College London. He also teaches research methodology and statistics, which shows throughout this book in the didactic approach he takes to describing the mathematical and philosophical basis of statistics. Specifically, the author seeks to explain, through this book, “how a test procedure works from the ground up” (p. xiii), something he notes to be missing from most statistics handbooks. The lack of patient and clear explanations – especially for students, but not only – regarding statistical thinking and ways of perceiving the basic principles of statistics is listed as an important motivation for the book. And while it contains important information for researchers in various fields in need of statistics, it specifically narrows down questions of experiment design and methodology for linguists (primarily working with corpora, but not only). In the preface to the book, the author briefly discusses the aspects that linguists need to know about statistics, and one point stands out: the bias that a framework inflicts onto our data and results. This launches the reader into the first chapter and the first taste of the book.

Structurally, the book is made up of six parts, each divided into multiple chapters, which discuss specific aspects for corpus linguistics research, as well as more general issues within the field of statistics, such as experiment design, confidence intervals, effect sizes, sampling, resampling, and more. The book also contains two appendices with practical knowledge, as well as

---

\* “Babeş-Bolyai” University, Doctoral School of Linguistic and Literary Studies, mihaelabuzec312@gmail.com.

a few supporting notes in navigating the book, including a glossary, an index, explanations for terminology and notations, and references for further reading. I will very briefly present each part of the book, with its main contributions.

Part 1 is called “Motivations” (pp. 1-24), and it contains only one chapter, “What might corpora tell us about language?”. This part provides examples of what sort of data can be used for different linguistic analyses; the importance of annotation, abstraction, and analysis of corpora for qualitative analysis, experimentation, and exploration; and the three types of empirical evidence linguists can obtain from corpora (or any data source, for that matter): factual evidence, frequency evidence, and interaction evidence. Also part of this chapter is a defense of corpus linguistics. Acknowledging the pertinent criticism (of, among others, Chomsky) that some researchers would only summarize facts and frequencies observed in the data without any significant reference, Willis argues that correct use of empirical evidence combined with reference to linguistic theory can offer invaluable insight into not only external manifestations of language, but also internal language and, therefore, psycholinguistics.

Part 2, “Designing experiments with corpora” (pp. 25-94), contains chapters 2 through 5, and discusses, at significant lengths, the issue of framework and baselines in the context of experiment design and data collection. This is an important part for linguists, as it tackles data gathering, hypotheses formulation and evaluation, going from a hypothesis to an experiment, performing the experiment and extracting meaningful data, and reporting results. The author also tackles methods and techniques which researchers can use, with examples and added explanations. Importantly, in the third chapter of the book, “That vexed problem of choice”, the ‘Per Million Words’ baseline is retrospectively evaluated and alternatives are proposed, and in the fifth chapter, “Balanced samples and imagined populations”, the problem of the representativeness of a corpus is discussed.

Part 3, “Confidence intervals and significance tests” (pp. 95-218), is another lengthy and heavy part of the book, separated into chapters 6 through 13. These tackle the issues of inferential statistics, probability, confidence intervals, variables, frequency evidence comparison, and replicating results. This part combines explanations of the mathematical and philosophical bases of statistics with actual concepts and formulae that need to be used, as well as examples of data sets and analyses that make it easier to understand and internalize such concepts. Chapters 8, “From intervals to tests” (pp. 134-165), and 13, “Choosing the right test” (pp. 205-217), are important resources for discriminating between the very large number of tests in statistics on the basis of the type of research being carried out. Chapter 12, “The replication crisis and new statistics” (pp. 178-193), ends with some recommendations and a checklist for empirical linguists, which feel personal and easy to apply, as though you were receiving feedback from a teacher. Among such complicated notions as those discussed in this part, these helpful tips might be easy to miss, but must not be.

Part 4, “Effect sizes and meta-tests” (pp. 219-260), is a continuation of Part 3 and considers more statistical methods. Among the important conceptual items related to statistical analysis, this part discusses significance tests, linguistic variables, how to compare results with other researchers using proper tests and methodology, and experiment design and refining.

Part 5, “Statistical solutions for corpus samples” (pp. 261-294), reviews the problems of imperfect data and the importance of correctly cleaning out your data. The author discusses the unreliability of some automated annotation algorithms and even sometimes of human correction. Different tests need to be carried out and analyses need to be reviewed in order to minimize bias or error.

Part 6, “Concluding remarks” (pp. 295-316), is made up of two chapters. Chapter 18, “Plotting the Wilson Distribution” (pp. 297-312), further considers crucial statistics concepts, offering an example of self-assessment and evaluation specifically tailored to plotting this type of distribution. Chapter 19, “In conclusion” (pp. 314-316), is the final one of the book; it offers the reader a word of encouragement, reiterating that, although hard work, statistical understanding is



paramount in becoming a good researcher. Many studies lose value because researchers do not understand how to craft a proper methodology, and the reported findings often end up being insignificant. One more aspect mentioned in this final chapter is the importance of reporting non-significant or negative results and findings, or even noticed mistakes in the methodology. As the author notes, “reporting defects explicitly may not be popular, but it is essential. [...] There is no shame in recognizing the limits of your data and methodology when you write up your research. On the contrary, it is only through honesty in reporting that science can progress” (p. 316).

The two appendixes that follow are: one presenting the Interval Equality Principle in detail, and the other containing samples of pseudo-code for computational procedures. The glossary at the end of the book is a particularly important resource for those just getting started with statistics and who might be lost in the many names and terms it operates with.

Of course, one needs to critically assess each part of the book and refer again and again to the type of study one carries out, since not all pieces of information will serve everyone the same. However, going through this book, it is important to keep a notebook near you and write down the questions you are being asked about your research, because it will help delimit your data sets better and, eventually, be more aware of what statistical information you need, and what you can leave out. Despite an intimidating and complicated topic, the book reads well. The didactic character of the author is seen in the way he frames problems in stories and actual experiments or research. Through this, Wallis brings statistics closer to the reader and emphasizes the importance of understanding the basic concepts and issues before venturing into the experiment proper or the analysis. A multitude of visuals help the reader see the data more easily, all while proving the importance of visuals and providing an example of well-designed graphs. The conclusions at the end of each chapter help the reader make sense of each part and more easily wrap one concept and move to the next.

The invitation of the book to always question your data and to always critically assess your methodology are important habits researchers need to develop, along with the classic ‘correlation is not causation’, but even beyond it. Other myths of statistics are discussed and alternatives are proposed, such as why “rich” data might be significantly more valuable than “big” data, and others that all in all help a person refine their research self.

Overall, the merit of this book is clear, and actually made explicit by the author: “demystify[ing] statistical reasoning in the conduct of corpus linguistics research” (p. 314). It is a book for students and researchers alike; there is no shame in going over the basics whenever starting a new research, because a fault in data sampling or in methodology design can snowball into the results and findings, and render them rather irrelevant. As such, this book should be a resource linguists – especially those working with corpora but not only – should return to again and again, until these principles of statistics become second nature.



*Reproducerea integrală sau parțială, multiplicarea prin orice mijloace și sub orice formă, cum ar fi xeroxarea, scanarea, transpunerea în format electronic sau audio, punerea la dispoziția publică, inclusiv prin internet sau prin rețele de calculatoare, stocarea permanentă sau temporară pe dispozitive sau sisteme cu posibilitatea recuperării informațiilor, cu scop comercial sau gratuit, precum și alte fapte similare săvârșite fără permisiunea scrisă a deținătorului copyright-ului reprezintă o încălcare a legislației cu privire la protecția proprietății intelectuale și se pedepsesc penal și/sau civil în conformitate cu legile în vigoare.*

---

© EUB-BUP pentru prezenta versiune

---

**<https://editura-unibuc.ro/>**

B-dul Mihail Kogălniceanu 36-46, Cămin A (curtea Facultății de Drept),  
Corp A, Intrarea A, etaj 2, Sector 5, București, România; tel.: + (4) 0726 390 815  
e-mail: [editura.unibuc@gmail.com](mailto:editura.unibuc@gmail.com)  
Librărie online: <https://editura-unibuc.ro/magazin/>  
Centru de vânzare: Bd. Regina Elisabeta, nr. 4-12, București, tel. + (4) 021 305 37 03

---

[tipografia\\_unibuc@yahoo.com](mailto:tipografia_unibuc@yahoo.com)

Bd. Iuliu Maniu 1-3, Complex LEU  
tel: 0799 210 566

Tiparul s-a executat sub cda 4917/2021  
la Tipografia **Editurii Universității din București – Bucharest University Press (EUB-BUP)**

---

